

# Automated Decision-Making Systems (ADS) Workgroup Meeting

---

September 23, 2021

# Agenda for September 23, 2021 Meeting

## Agenda

- 2:30 Welcome and administrative updates – Katy Ruckle
- 2:40 Washington One System Review – DOC & Workgroup
- 3:40 Discussion of what (if any) ADS uses should be prohibited? – Jon Pincus
- 4:00 Workgroup Discussion – All members
- 4:10 Answers to Open Tasks – Assigned work group members
- 4:20 Open Discussion
- 4:30 Adjourn

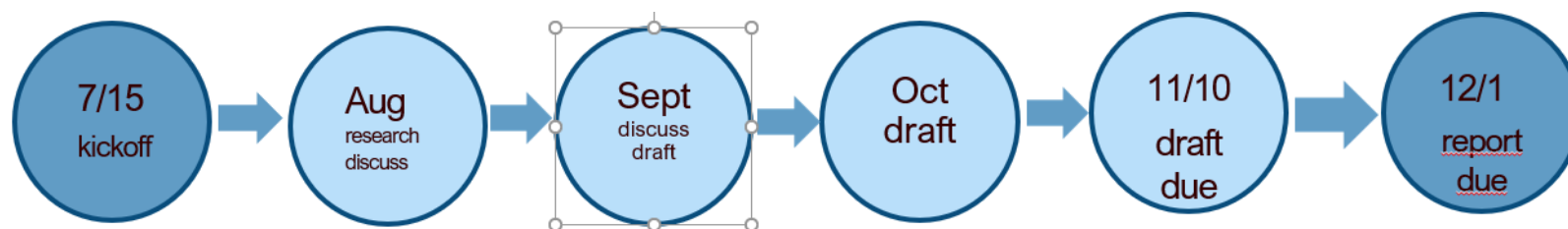
O  
P  
D  
P

# Welcome and Administrative Updates

O  
P  
D  
P

# Administrative Updates

- Teams Channel is not working for external members. I can try to continue trouble shooting, but easiest to just work with whole group by email at this time.
- Timeline



**Automated  
Decision-making  
Systems Workgroup**

O  
P  
D  
P

# Washington One System Review

O  
P  
D  
P

## Automated Decision-Making System Questions: Review of Washington One

### 1. What is the automated decision system's name, vendor and version?

Washington One

Vendor – procured in 2015; Dr. Zachary Hamilton and WSU

Ongoing maintenance and updates performed in-house at DOC.

**#1: What was the level of involvement of DOC in the design process of the decision system?**

O  
P  
D  
P



## Automated Decision-Making System Questions: Review of Washington One

### 2. What decisions is the system utilized to make?

Identifies risk level classifications for all individuals under DOC jurisdiction. Not just those housed at DOC facilities.

Program priorities are identified – what an individual's domain needs are. (i.e. high substance abuse need prioritized for substance abuse treatment)

Used to make determinations concerning risk-based supervision

(The output need domains used for decision making: Aggression, Attitudes/Behaviors, Education/Vocational, Employment, Mental Health, Residential, Social Influences, Substance Abuse) – High/Mod/Low score

#### *2. used to make determinations concerning risk based supervision.*

- I'd like to know what the implications of the risk score is for individuals? Does this impact their ability to be released? Does a high risk score make it more likely that they would be re-incarcerated?
- Risk score is used to determine level of supervision in the community and program priority in facilities, it does not impact their ability to be released. A higher risk score indicates the likelihood that an individual will recidivate compared to others under DOC jurisdiction. However, it does indicate outcomes. Meaning, just because someone is higher risk does not mean they will recidivate.

## Automated Decision-Making System Questions: Review of Washington One

### 3. What data is inputted into the system?

- a. Approximately dynamic manually inputted responses as well as data from system (106 items, 20 static, 9 semi-dynamic, 77 dynamic)
- b. Questions on the following domains: Demographics, Juvenile Record, Adult criminal history, correctional events, residential status, education, vocational(work) status, employment status, relationship status, family status, children, friends, leisure time, alcohol/drug use, mental health (including hospitalization & medications), aggression, attitudes/behaviors
- c. Criminal History + Needs Interview + Correctional Events = Washington One Recidivism Model

### 3. need definitions for dynamic and static data.

#### b. demographics? Why is this used?

#### How is mental health assessed? Is criminal history adjusted for potential racial bias in arrests?

- Demographics: Age, Gender, Education level, Employment, Income, Martial status, Children. These are collected as they are indicators of risk, need, and potential recidivism. (Note: racial/ethnic demographics are not part of the tool and are not collected)
- Mental health: Documented mental health diagnoses, mental health problem indicated from file or individual self-report, suicidal thoughts/attempts, hospitalization/in-patient stays for mental health, out-patient mental health treatment, mental health medication prescribed
- No criminal history is not adjusted for potential racial bias in arrests at this time. Criminal history does not include arrests, only convictions.



**4. How is the input data gathered, how often is it updated, and are subjective inputs ever audited for consistency across data collectors?**

- a. Mix for input data gathered. OMNI – Offender Management Network of Information.
- b. Gathered by counselors via interview with the individual. “Motivational interview” to elicit responses from the person being interviewed. Conversational – to gather narrative and context.
- c. Reassessment schedule – when the conviction record is updated. But can occur at anytime based on new information. At intervals based upon time to Eligible Release Date (ERD) or Supervision End Date (SED)
  - i. Criminal Conviction Record update (within 30 days)
  - ii. Every 6 months to a year based on ERD/SED
  - iii. Within 30 days before transfer to partial confinement
  - iv. Between 60 and 90 days before release to the community from a work/training release
  - v. When required as part of a quality of review process
- d. Inter-rater reliability not automated at this time re subjective inputs.
- e. Audited initially during development.
- f. There is a QA performed on assessments – not accuracy across. Supervisor review for accuracy.
- g. Reviews done on all pre-closure cases.
- h. Released – low or moderate audit process performed by correction specialist.

## Automated Decision-Making System Questions: Review of Washington One Questions

### 4. not sure what OMNI is? Are counselor's assessments checked for bias?

#### d. not sure what "inter-rater reliability" is.

- Offender Management Network Information, it is DOC's case management tool
- Inter-rater reliability is in brief: if two people were to both conduct the assessment would they come to the same conclusion? Testing for inter-rater reliability is testing for bias in counselor's assessments. The tool has been tested for inter-rater reliability but is not ongoing at this time.

**Question 4:** Earlier, it was mentioned that interviews are conducted with neighbors, friends, family, etc. as part of the data gathering process. Is that accurate? If so, what are the questions asked of these individuals?

I don't believe that's accurate. The assessment is scored using results from the interview process and documented file material. Interviewing collateral contacts is not part of the process. However, a case manager may select item responses based upon their knowledge of the individual.

on #4, how does the "Motivational interview" relate to the "High Risk Situation interview" referred to in publication 400-BR011?

- Unable to locate this publication before meeting, but to clarity, DOC does not specifically use motivational interviewing, the process is similar to motivational interviewing.

## Automated Decision-Making System Questions: Review of Washington One

### 5. Is the decision algorithm available for examination by the agency and/or the public?

- a. Yes for the agency to a certain extent, but not the public.

**Question 5:** What do you mean by responding that the decision algorithm is available to the agency “to a certain extent”? - #5 says the algorithm is available to the agency "to a certain extent". Could you go into more detail on the boundaries and implications of this?

DOC has the weights. What we don't have is full methodology behind how the weights were created.

### 6. Has there been any public or community engagement used in selection or design of the system? If so, please describe this engagement.

- a. There was a bid process that was posted publicly.
- b. There has been engagement with the Family Council about the Washington One system to answer questions about the assessment.
- c. Pre-implementation presentations and consultations with AAG

### 7. Does law or regulation mandate any of the decision system criteria? If so, which criteria?

Required to use the risk assessment system endorsed by WSIPP **see RCW [72.09.270](#)**

Criminal justice system –  
WSIPP

Levels of risk is relevant.

## Automated Decision-Making System Questions: Review of Washington One

### **8. Do the system's decisions intentionally differentially affect members of protected classes, such as selecting persons with disabilities for certain benefits?**

- a. Yes, differentiates by gender to allow tool to assign supervision levels. System is designed to be gender responsive to consider differential needs based on gender in the algorithm. Intended to provide for needs of the group by also identifying case management relevant issues such as ADA, behavioral health, medical and other needs that may indicate a need for supportive activities

### 8. does gender include categories beyond male and female?

- Not at this time

### **Question 8: What is the basis for determining that there are differential needs based on gender? What about individuals who are non-binary?**

A large amount of research on the RNR model and gender (Andrews, Bonta, Salisbury, Van Voorhis, Chesney-Lind and more). The tool is only male/female at this time. A non-binary individual would (and I would have to check policy) be assessed using their assigned gender at birth. Research on risk assessments and the RNR model for non-binary and transgender individuals is in its infancy. Because these individuals make up a smaller overall portion of the incarcerated/supervision population (In WA it's less than 1%) it is hard to conduct statistically rigorous analysis and modeling.

. #8: In this answer you mention things such as "[s]ystem is designed to be gender responsive" or "intended to provide for needs of the group." Is there a public official document where this type of design choices and purposes are recorded?



## Automated Decision-Making System Questions: Review of Washington One

### 9. Has the system has been tested for unintended bias by the agency or an independent third party? If so, what were the results? Describe briefly the nature of the testing.

- a. Yes, during system's development it went through initial analyses of biases that were considered in design.
- b. WSIPP contract to evaluate current state of potential of bias based on demographic.
- c. Another analyses of bias post-*Blake* decision current state.
- d. Norming period after tool launch considering potential bias and making adjustments.

**Question 9:** What were the results of the testing for unintended bias? On this subject, we find it not very credible that a criminal justice system could be designed without any unintended racial biases. If the DOC is consistently finding none, that leads us to question the depth of the examination. Additionally, did WSIPP have access to the full algorithm and all of the proprietary weights?

We are currently having the tool assessed by a third party. DOC has not made claims on this matter as is waiting for the third party review. Yes WSIPP has the weights.

on #9, Did the limitations on visibility to the algorithm mentioned in #5 have any impacts here? Did they apply to WSIPP as well? Are the processes and results that were used for any of these tests public?

No. Processes and results from WSIPP reports and DOC reports are public, there are also published articles on the development of the STRONG-R. (When I say processes I mean that the methodology is discussed. The step by step detailed process and/or code is not public).



# Automated Decision-Making System Questions: Review of Washington One

also on #9 a wording question. "Algorithmic audits" are something of a term of art. I wonder whether it would be useful to reflect that in the question, maybe with wording like "Has the system been tested or audited for unintended bias?" Would that have changed your answer or the process of getting the information? Other thoughts on this?

O  
P  
D  
P

## Automated Decision-Making System Questions: Review of Washington One

### **10. Has the system has produced known erroneous results and if so, what were those errors (including the results of any audits conducted to check for erroneous results)?**

- a. No known erroneous results or major audits beyond QA work.
- b. Known data entry issues which are identified and corrected through QA process.
- c. Norming period report in Oct. 2020. Published by WSIPP.
- d. System went live Dec. 15, 2017.

**Question 10:** The response states that there are no known erroneous results. Again, given a system of this complexity, this leads to a questioning of the thoroughness of the examination. What did the QA process entail (assuming QA means quality assurance), and if there were no major audits, how can DOC be sure that there were no erroneous results?

Difference between erroneous results in the tool versus data entry. Yes there have been data entry errors causing incorrect scores but to our knowledge the tool itself has not produced an erroneous result when presented with the correct inputted information.

- on #10: is there anything public on the QA process? The norming report notes that there were minimal differences between contact requirements on the previous system and the Washington ONE system; did QA in general take a similar approach of looking for differences with the previous system?

Yes, the pilot study report discusses the QA process and is public.

**11. In addition to any intentional differential effect on members of a protected class, are there are other differential effects on protected classes as shown by comparison of the system's data to general census data or, where relevant, subpopulation data, such as the effect on justice system defendants of color as contrasted with all defendants? If audits have been performed to determine such differential effects, what were the results of those audits?**

- Yes, some differential effects are expected (especially for the needs domains which determine treatment planning) depending on the protected class. E.g., someone with a history of mental and behavioral health issues/diagnoses may score as high needs in Mental Health. Mental/behavioral health issues is a high correlate with many co-occurring disorders/disabilities)
- Census data as a benchmark for fairness or bias is unreliable.
- Other differential effects are due to criminal justice system as a whole - e.g., Age curves.
- Same response as number 9.
- DOC contracting with third party to analyze again and in the future.

**Question 11:** The response to question 11 does not answer the question of whether there are differential effects on populations of color. See comments above. Race is not considered in the tool.

## Automated Decision-Making System Questions: Review of Washington One

- #11: great to see that there are ongoing contracts to study biases. This is certainly something we should recommend as best practices. I have to say though that given the biases reported in risk assessment systems elsewhere, it's very surprising that there haven't been at least some unintentional biases detected in this system to date ... quite frankly it makes me wonder how effective the analyses have been (and I don't mean this with any disrespect to the teams doing the analysis, the state of the art has advanced a lot in the last few years).

There has not been an in depth examination of racial biases in the Washington ONE to date, it is currently in process. The tool was launched in December of 2017, the discussion of the WSIPP assessment began in spring of 2021 following the norming period study. There has to be some passage of time to collect data to accurately assess the tool for biases.

O  
P  
D  
P



## Automated Decision-Making System Questions: Review of Washington One

### **12. Can those affected by a system decision review and challenge the basis for that decision? If so, how, and what were the results of any such challenges?**

- a. Yes, individuals can challenge the results of an assessment.
- b. It is in DOC policy.
- c. Process goes to superintendent or field administrator.

**Question 12:** This does not answer (or link to) how individuals may review and challenge decisions and what the results of any challenges were. DOC 320.400 states that, "Individuals may appeal assessment results in writing to the Superintendent/Work Release Administrator/Field Administrator or their designees within 48 hours of receiving the results of an RNR assessment. The decision made by the Superintendent/Work Release Administrator/Field Administrator or their designees is final."

**#12:** Has there been any judicial litigation against a system decision (I am thinking about tort law or liability in discrimination law)? If so, has the court had the chance to access the decision algorithm for examination?



## Automated Decision-Making System Questions: Review of Washington One

**12. Can those affected by a system decision review and challenge the basis for that decision? If so, how, and what were the results of any such challenges?**

- a. Yes, individuals can challenge the results of an assessment.
- b. It is in DOC policy.
- c. Process goes to superintendent or field administrator.

**13. Is the decision system is operated by a third party? If so, what rules govern such operation and what audits are conducted to ensure compliance?**

- a. No, system is not operated by a third party.

**14. What is the fiscal impact of the system, including initial cost, operating costs, and any cost savings established as flowing from use of the system?**

- a. All costs not readily available. RFP – 2015 Work started in 2016.
- b. Mandated by law to have the system.

- #14: is there a better way of framing the question so that it could be answered at least partially? For example the ongoing operating costs (internal staffing, vendor payments if any, contracts to study bias) are presumably knowable. And the norming report notes that it didn't lead to changes in contact workload, which arguably relates to this general topic.

## Automated Decision-Making System Questions: Review of Washington One

### **15. What were the personnel hours required to gather the relevant information (questions 1-14) for the system examined?**

a. 15.5 hours

- #15: it would be great to hear about roughly where the time went and if there are any obvious ways to streamline it ... quite possibly no but it can't hurt to ask.

O  
P  
D  
P

## Proviso details

Are there changes needed regarding the development, procurement, and use of ADS by state agencies?

If yes, what types of changes?

Development?

Procurement?

Use?

How can ADS be reviewed before adoption?

How can ADS be reviewed while in operation?

Audited to ensure ADS is fair, transparent, and accountable?

How can state ensure ADS does not improperly advantage or disadvantage particular residents?

O  
P  
D  
P

Discussion of what (if any) ADS uses should be prohibited?

O  
P  
D  
P

# When should agency uses of ADS and AI profiling systems be prohibited?

ADS Workgroup meeting, 9/24

[Jon Pincus](#)

jon@achangeiscoming.net



# Proviso language

Workgroup must examine

- (i) When state agency use of automated decision making systems should be prohibited;
- (ii) When state agency use of artificial intelligence-enabled profiling systems should be prohibited

# Broad trends

## Reasons for prohibitions

- “Unacceptable risks”
- Impact on rights
- Discrimination

## Prohibitions to date focus primarily on

- Facial recognition and other biometrics in public spaces, \
- Predictive policing

## Now seeing more proposals for prohibitions on

- “Social scoring”
- Emotion manipulation and tracking
- Categorizing people into clusters on protected classes

# Examples of prohibitions to date

- [King County becomes first U.S. county to ban government use of facial recognition software](#) (*Geekwire*, 6/21)
- [Vermont lawmakers approve ban on facial recognition technology](#) (WCAX, 10/20)
- [Maine Now Has the Toughest Facial Recognition Restrictions in the U.S.](#) (Slate, 7/21)
- [Santa Cruz bans predictive policing in U.S. first](#) (*Reuters*, 6/20)
- [Oakland To Ban Predictive Policing, Biometric Surveillance Tech](#) (*SF Gate*, 1/21)

[Algorithmic Accountability for the Public Sector](#)'s summary (Ada Lovelace Institute, AI Now, Open Government Partnership, August 2021)

“Some jurisdictions have banned or prohibited the use of particular kinds of ‘high risk’ algorithmic systems.... Prohibitions and moratoria have been most prominently applied to facial recognition technologies used by law enforcement, and in some cases local governments.”

# EU Draft AI directive (April 2021)

## Prohibits systems with “unacceptable risk”

- Manipulative systems likely to cause harm that use subliminal techniques or exploiting vulnerabilities due to age or disability
- “Social scoring”: AI systems i) used by or on behalf of public authorities, ii) to generate ‘trustworthiness’ scores and which ii) lead to either unjustified or disproportionate treatment of individuals or groups, or detrimental treatment which, while justifiable and proportionate, occurs in an unrelated ‘context’ from the input data.
- Some uses of real-time biometric systems (including facial recognition) in publicly accessible spaces by law enforcement, with exceptions

## Discussion in

- Khari Johnson's [The Fight to Define When AI Is ‘High Risk’](#) in *Wired*
- Michael Veale and Frederik Zuiderveen Borgesius, [Demystifying the Draft EU Artificial Intelligence Act](#)

# Proposals to strengthen EU Draft

[European Data Protection Board](#): “intrusive forms of AI – especially those who may affect human dignity – are to be seen as prohibited”

- Automated recognition of human features in publicly accessible spaces
- Any type of social scoring
- Categorizing individuals from biometrics into clusters according to ethnicity, gender, as well as political or sexual orientation
- Inferring emotions

[European Digital Rights + civil society orgs](#): limit deployments of AI that “unduly restrict human rights”

- Biometric mass surveillance
- Uses of AI at the border and in migration control
- Social scoring and AI systems determining access to social rights and benefits
- Predictive policing
- Use of risk assessment tools in the criminal justice system and pre-trial context



## UN Human Rights Commission report (September 2021)

The higher the risk for human rights, the stricter the legal requirements for the use of AI technology should be

- A risk-proportionate approach to legislation and regulation will require the prohibition of certain AI technologies, applications or use cases, where they would create potential or actual impacts that are not justified under international human rights law, including those that fail the necessity and proportionality tests.
- Moreover, uses of AI that inherently conflict with the prohibition of discrimination should not be allowed. For example, social scoring of individuals by Governments or AI systems that categorize individuals into clusters on prohibited discriminatory grounds should be banned in line with these principles.
- States should also impose moratoriums on the use of potentially high-risk technology, such as remote real-time facial recognition, until it is ensured that their use cannot violate human rights.

# Are prohibitions on “fully automated” systems useful?

Ben Green in [The Flaws of Policies Requiring Human Oversight of Government Algorithms](#)

- Evidence suggests that people are unable to perform the desired oversight functions.
- Human oversight policies legitimize government use of flawed and controversial algorithms without addressing the fundamental issues with these tools.
- Thus, rather than protect against the potential harms of algorithmic decision-making in government, human oversight policies provide a false sense of security in adopting algorithms and enable vendors and agencies to shirk accountability for algorithmic harms.

# Discuss!

- (i) When state agency use of automated decision making systems should be prohibited;
- (ii) When state agency use of artificial intelligence-enabled profiling systems should be prohibited

# Workgroup Discussion

O  
P  
D  
P

# Definition of ADS in Proviso

(f) For purposes of this subsection, "automated decision system" or "system" means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that

- uses data-based analysis or calculations to make or support
  - government decisions,
  - judgments, or
  - conclusions
    - that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation:
      - benefits,
      - protections,
      - required payments,
      - penalties,
      - regulations,
      - timing,
      - application, or
      - process requirements.

## Definitions from Task 5.01

The definition of ADS below captures algorithms, including rules-based and machine learning algorithms, that make or inform decisions that cause differential impacts on Washington residents. We believe that a definition that captures many types of ADS is important as both rules-based and machine learning algorithms have and can cause harm to individuals and communities. Ensuring that such algorithms are captured by the definition of ADS ensures that they may be examined for accuracy and bias.

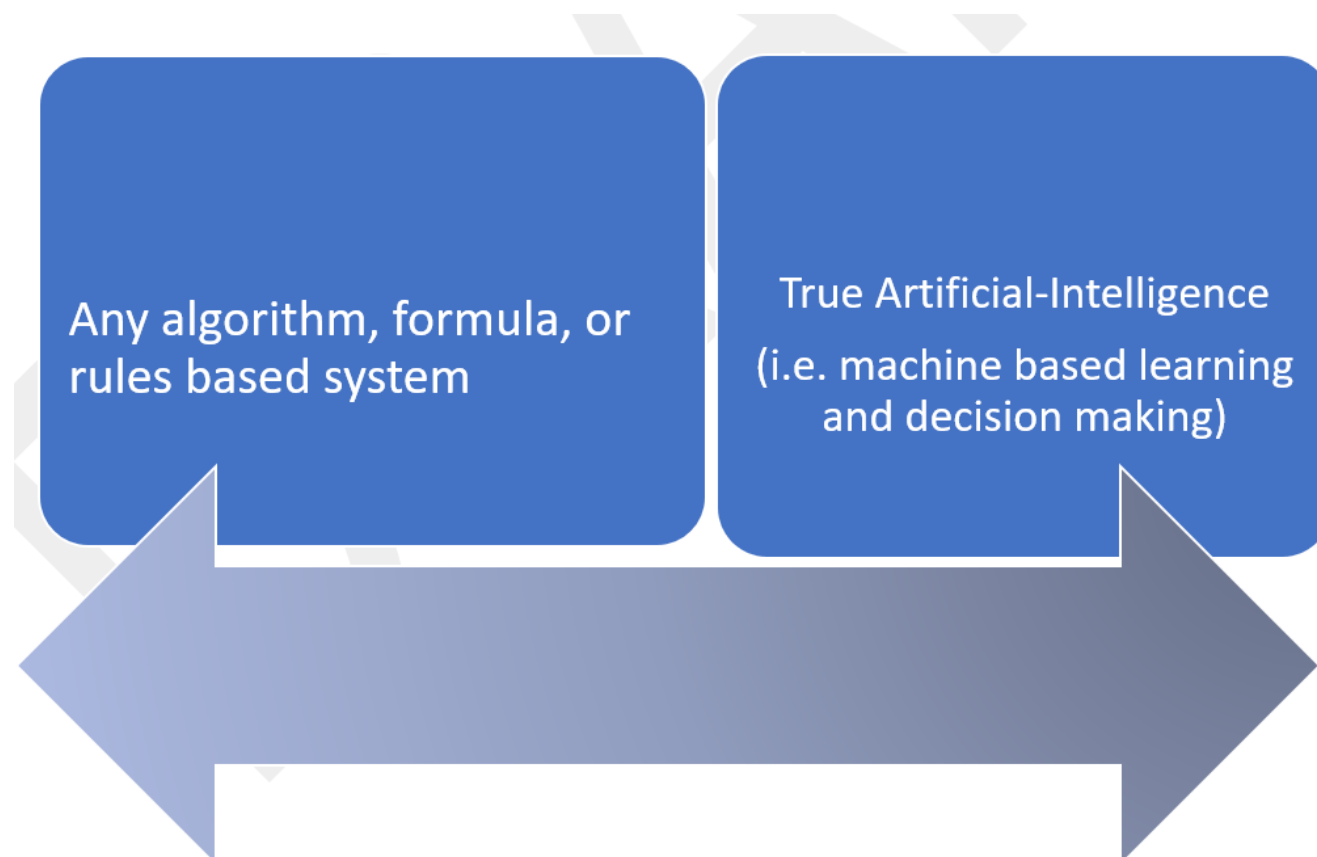
**"Automated decision system"** means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that uses data-based analysis or calculations to make or support government decisions, judgments, or conclusions that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation, benefits, protections, required payments, penalties, regulations, timing, application, or process requirements.

**"Automated decision system"** does not include tools that do not make or support governmental decisions, judgments, or conclusions that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of government interaction with that individual including, without limitation, internal governmental computer server or electrical usage optimization, antivirus programs, and internal governmental space optimization programs.

**"Automated final decision system"** means an automated decision system that makes final decisions, judgments, or conclusions without human intervention.

**"Automated support decision system"** means an automated decision system that provides information to inform the final decision, judgment, or conclusion of a human decision maker.

# Spectrum on Scope and Definition of ADS



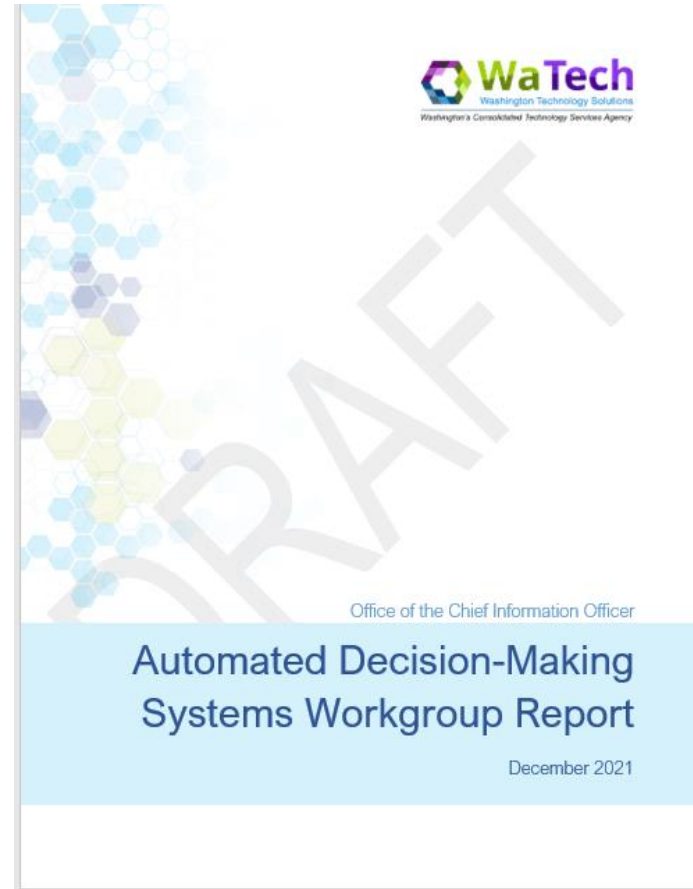
O  
P  
D  
P



# Report Structure and Policy Recommendations

O  
P  
D  
P

# Report Drafting



O  
P  
D  
P

## Answers to Open Tasks – Assigned work group members

O  
P  
D  
P

Action Item*	Description	Person Responsible	Resolution
05.01	Provide the definitions that are a closer match to SB 5116 than those provided in the budget proviso.	Jen Lee	Completed 9-9-21. See Definitions in slide deck.
05.02	Start a draft report outline and share the document in Teams.	Katy Ruckle	09/23/21 – To be completed by 10-4-21
05.03	Submit their agency/organization logo to Katy Ruckle.	All Workgroup Members	10/07/21
05.04	Conduct outreach to identify any systems an agency would be willing to review for the report.	State Agency Workgroup Members	10/07/21
05.05	Identify and share resources equity-based audits from the Poverty Reduction Workgroup.	Nancy Aguilar	10-7-21

# Open Discussion

O  
P  
D  
P

# Thank you!

O  
P  
D  
P